Development of a Fluorescent Reporter-System to Quantify Gene Expression *in vivo* in the Hyperthermophilic Archaeon, *Thermococcus kodakarensis*

Research Thesis

Presented in partial fulfillment of the requirements for graduation *with research distinction* in the undergraduate colleges of The Ohio State University

by Ellen Zwick

The Ohio State University May 2015

Project Advisors: Dr. John N. Reeve, Department of Microbiology

Dr. Charles J. Daniels, Department of Microbiology

The biological world is divided into three Domains; the Archaea, Bacteria and Eukaryotes (1, 2). The Archaea and Bacteria are similar in morphology and gene organization. They are also both single-celled prokaryotes, but they are evolutionarily distinct, with the Archaea more closely related to the Eukaryotes than are Bacteria (1, 2, 3). Consistent with this, the Archaea and Eukaryotes have many molecular features in common that differ from their bacterial counterparts, most notably transcriptional and translational machinery (3, 4). With this in mind, the Archaea have been exploited as relatively simple model systems to study DNA replication, transcription (RNA synthesis) and translation (protein synthesis) (5, 6, 7, 8, 9). In addition, as many Archaea inhabit extreme environments, they are also attractive research systems to investigate extremophile-life and as sources of unusually stable enzymes that may be developed as catalysts in industrial biotechnologies for "green" chemistry (10, 11).

Research on the extremophile Archaea has, however, been limited by the difficulty of their laboratory cultivation and, before the discovery in 2003 that *Thermococcus kodakarensis* is naturally competent for DNA uptake and *in vivo* recombination (12), by a lack of available genetics tools. It is a strictly anaerobic Archaeon that grows optimally at 85°C. *T. kodakarensis* was isolated from solfatara on the shore of Kodakara Island in Kasoshima, Japan (13). Many hyperthermophilic related members of the Thermococcales genus have since been isolated from a variety of geothermal environments (14). The Thermococcales are abundant and apparently contribute substantially to the ecology and metabolic activities within many high temperature ecosystems (15). Based on their genome sequences, they also have many molecular biology features in common with Eukaryotes (3, 4). Research investigations of

the Thermococcales will add to knowledge of hyperthermophilicity, metabolism in extremophile environments and to basic gene expression mechanisms in both the archaeal and eukaryotic domains.

Since the discovery that *T. kodakarensis* can be genetically manipulated, this species has been developed as the predominant model system for archaeal research (14). Plasmids, designated shuttle vectors, have been constructed that replicate and express genes in both *E. coli* and *T. kodakarensis* (16, 17). Dr. Reeve's research group is currently deleting every non-essential gene individually from the *T. kodakarensis* genome to establish an approximately 2,300-strain gene library. However, a more in-depth understanding of the archaeal genome necessitates more complex genetic tools, particularly for a simple reporter-system. Reporter gene systems are integral tools in molecular biology to detect and quantify gene function and regulation. There is a reporter system in place for *T. kodakarensis,* however, it is laborious and its function is limited (17). Furthermore, most commonly used reporter systems employ a fluorescent or luminous protein, such as green fluorescent protein (GFP), to report the strength of gene expression. These proteins typically originate from mesophilic species and, unfortunately, they are unstable at high temperatures. In addition, they require oxygen to function, and so cannot be used as reporters for gene expression in any hyperthermophile and/or anaerobe, including *T. kodakarensis.* The goal of my project is to construct a reporter system that is thermostable and functions in anaerobically growing *T. kodakarensis*.

Background

The goal of this project is to develop a versatile reporter system for *T.*

kodakarensis patterned on a reporter system developed and commercialized by New England BioLabs (NEB) called the SNAP-tag system (18). This reporter system employs a derivative of a human DNA-repair enzyme, O^6- methylguanine-DNA methyltransferase (MGMT). MGMT belongs to the family of O^6 alkylguanine alkyltransferases, and reacts with O^6-methylated guanine nucleotides to remove the methyl group at the O^6 position and release the repaired guanine nucleotide back into the system (19). In this reaction, the methyl group is transferred and permanently attached to the MGMT protein, thus inactivating MGMT's enzymatic activity (19). O^6-methylated guanine nucleotides pair erroneously with thymine nucleotides, causing a G:C → A:T transitional mutation in DNA.

Multiple methylguanine analogs are commercially available from NEB, making this MGMT- based reporter system highly versatile. Amongst other analogs like biotinylated methylguanine, a broadly used affinity tag in molecular biology, this includes a variety of fluorescent probes with different photochemical properties. When supplied to cells synthesizing the MGMT protein, these analogues are bound by MGMT and permanently attached to the MGMT protein, thus making the MGMT protein quantifiable by its fluorescence. By standard genetic engineering, the gene encoding MGMT can be positioned as needed downstream of any expression genetic element being assayed, or can be fused in-frame to any protein whose expression is being directly investigated.

Project Objectives

The *T. kodakarensis* genome has a gene, designated TK1971, which encodes an

enzyme very closely related to the human MGMT protein (figure 1). The goal of this project is to develop this enzyme, designated TK-MGMT, and demonstrate its functionality as a versatile and thermostable reporter for use in *T. kodakarensis*. Based on the homology between the human enzyme and TK-MGMT, it was expected that TK-MGMT would also bind the commercially available reporter substrates from NEB, including methylguanine analogs with fluorophore at the O^6 position. The available fluorophores do not require oxygen to fluoresce and are chemically stable at 85°C, the *T. kodakarensis* optimal growth temperatures. When established, the TK-MGMT reporter system will be available to archaeal researchers to measure archaeal promoter strengths, to quantify protein expression, and to observe protein localization *in vivo* within *Thermococcales* cells.

```
Introduced muts                                IIFLG
hAGT_SNAP-basis      ---MDKDCEMERTTLDSPLGKLELSGCEQGLHEIKLLGKGTSAADAVEVPAPAAVLGGPE 57
TK-MGMT_TK1971_      MLSVEKFRVGERVVWIGVIFSGRVQGIAFAFDRGTLMKRIHDLAEHLGKRGVSISLD--- 57
                        ::*    **..  . : . .:.*   .:.. .*: :  . *: :   . :  *.
Introduced muts          A                                                  SH
hAGT_SNAP-basis      PLMQCTAWLNAYFHQPEAIEEFPVPALHHPVFQQESFTRQVLWKLLKVVKFGEVISYQQL 117
TK-MGMT_TK1971_      --VQPSDYPEKVFKVLIGELDNASFLRELSFEGVTPFEKKVYEWLTKNVKRGSVITYGDL 115
                       :* : : :  *:   .  : .    . ..    .* ::*   * * ** *.**:* :*
Introduced muts             AATA  KT LS             GGDLD  G EW
hAGT_SNAP-basis      AALAGNPKAARAVGGAMRGNPVPILIPCHRVVCSSGAVGNYSGGLAVKEWLLAHEGHRLG 177
TK-MGMT_TK1971_      AKALN--TSPRAVGGAMKRNPYPIVVPCHRVVAHDG-IGYYSSGIEEKKFLLEIEGVKEW 172
                     *   .  .:.*******: ** **::******. .* :* **.*:  *::**  ** :
                                  Δ
hAGT_SNAP-basis      KPGLGGSSGLAGAWLKGAGATSGSPPAGRN 207
TK-MGMT_TK1971_      TS---------------------------- 174
```

Figure 1. Clustalw alignment comparing the TK-MGMT protein (TK-MGMT_TK1971_) with the human MGMT protein (hAGT_SNAP-basis) utilized by NEB for the SNAP-tag system. The mutations introduced into the TK-MGMT protein (Introduced muts) are highlighted above the human MGMT amino acid sequence. Mutations that increased hHGT activity with benzylguanine (27) are highlighted in yellow. Mutations that further increase hHGT activity with benzylguanine (28) are highlighted in blue. Mutations that decrease the enzyme's affinity for DNA (29, 30) are highlighted in pink. Amino acids highlighted in gray indicate amino acid changes that decreased wild-type human MGMT's sensitivity to N9-benzylguanine derivatives. A Proline essential for protein activity at the MGMT catalytic site is highlighted in red. Amino acids in bold indicate amino acids not found in the TK-MGMT protein, but are present in the human MGMT protein.

The crystal structure of the Tk-MGMT protein is available, and has been demonstrated to be significantly more thermostable than its mesophilic counterpart due to

the intra-helical ion-pairs on the interior of its structure (21, 22). The MGMT protein has been demonstrated to have an important role in DNA repair in thermophilic archaea *Solfolobus solfataricus*, as well as fluorescently active with the same methylated guanine analogs utilized by NEB for the SNAP-tag reporter system *in vivo* in *E. coli* (23, 24).

Materials and Methods

Southern blot analysis

Southern blot analysis was carried out according the protocol outlined in Santangelo et al. 2007. Genomic DNA from *T. kodakarensis* strains TS559 and DJA7 was digested with New England BioLabs restriction enzymes *EcoRV* and *XhoI* and incubated with various DNA probes specific to the region surrounding the TK1971gene (figure 3). Bands were detected using a chromogenic detection method, using a Thermo ScientificTM PierceTM ECL 2 Western Blotting Substrate according to manufacturer instructions.

TK1971 Mutation Introduction

The mutations outlined in figure 1 were introduced on a pUC19-based plasmid containing the TK1971 construct (pDJ25, figure 4) using a QuikChange Lightning kit according to manufacturer instructions, or an inverse PCR strategy. The inverse PCR strategy utilized primers with a 3' overhang containing the desired nucleotide changes to amplify the plasmid, resulting in a linear product in which the mutated TK1971 gene was split into halves on either end of the pUC19 plasmid. This inverse PCR strategy also introduced an N-terminal *NotI* and C-terminal *SalI* restriction site on the TK1971 gene. The PCR product was ligated back together to create a circular plasmid with intact,

mutated TK1971. The mutations listed in figure 1 were introduced both singly and in combination into TK1971. The presence of the desired mutations in TK1971 was confirmed with sequencing.

The mutated versions of TK1971 were cut out of the pUC19 backbone using NEB restriction enzymes *NotI/SalI* and ligated into shuttle vector pTS543, which is capable of replicating in both *E. coli* and *T. kodakarensis*. Restriction enzyme digests were carried out according to NEB protocol. Ligation was carried out using T4 DNA Ligase, according to NEB protocol.

The pTS543 shuttle vectors with TK1971 inserts were then transformed into *T. kodakarensis* strain DJA7 according to transformation protocol outlined in Santangelo et al. 2010. Successful transformants were confirmed with PCR amplification off the pTS543 plasmid. Once successfully transformed into *T. koakarensis,* strength of TK1971 expression as well as its ability to remove and bind tagged methyl groups from methylguanine analogs was measured using confocal scanning laser microscopy (CSLM).

Strains		
TS559	ΔpyrF; ΔtrpE::pyrF; ΔTK0664, ΔTK0149	Santangelo et al. 2011
DJA7	TS559 with ΔTK1971	This study
DJA8	DJA7 with pDJ25	This study
DJA9	DJA7 with pDJ66	This study

Table 1. Table detailing the strains of *T. kodakarensis* used in this project.

Confocal Scanning Laser Microscopy

Samples of *T. kodakarensis* for CSLM were prepared and images were taken by Dr. Dominik Jäger. To prepare samples of *T. kodakarensis* cells for observation with

CSLM, *T. kodakarensis* was grown overnight in 20mL of rich media (ASW-YT+Pyr or ASW-YT+S medium). ASW-YT+Pyr medium was composed of artificial seawater (0.8× ASW), yeast extract (5.0 g/L), tryptone (5.0 g/L), and pyruvate (5.0g/L). ASW-YT+S^0 medium was composed of artificial seawater (0.8× ASW), yeast extract (5.0 g/L), tryptone (5.0 g/L), and sulfur (2.0g/L). Though both ASW-YT+Pyr and ASW-YT+S^0 media can be used, ASW-YT+Pyr was preferred because the ASW-YT+S^0 media required the elemental sulfur to be filtered out during sample preparation. The OD_{600} of the cultures was measured using a Bio-Rad spectrophotometer. The cultures were brought into am anaerobic chamber and either 0.1 or 0.2 OD_{600} of cells were spun down at 6000rpm for 2 minutes. The supernatant was removed and the cell pellet was re-suspended in labeling media and transferred to an amber-orange 1.7uL microcentrifuge tube to limit the media's exposure to light. The composition of the media depended on the OD_{600} of *T. kodakarensis* cells spun down. For 0.2 OD_{600}, the cells were re-suspended in 1.0ul of SNAP-Cell TMR-Star with 1.0ul BSA and 198uL complemented ASW-YT+Pyr. For 0.1 OD_{600}, the cells were re-suspended in 0.5ul of SNAP-Cell TMR-Star with 1.0ul BSA and 198.5uL complemented ASW-YT+Pyr. The final concentration of SNAP-Cell TMR-Star in both compositions was 5 μM.

The re-suspended samples were incubated for 1 hour at 85°C, and then spun down again at 6000rpm for 5 minutes. The supernatant was removed and the cells were re-suspended in 1.0mL of complemented ASW-YT+Pyr, spun down again at 6000rpm for 2 minutes and re-suspended in 1.0mL complemented ASW-YT+Pyr. This step was repeated five times to wash the cells. Two drops of NucBlue Live Ready-Probes Reagent (Hoechst 33342 DNA stain) were added to each resuspension and the cells were

incubated for 30 minutes at 85°C, spun down, and re-washed twice as previously described. The cell pellets were re-suspended in 50-100uL ASW-YT+Pyr or 150-200uL ASW-YT+Pyr for 0.1 and 0.2 OD_{600} of cells, respectively. These cells were kept in an anaerobic chamber on ice until they were prepared on microscope slides for CSLM.

Polylysine microscope slides were prepared in a darkened room under aerobic conditions. Samples were kept on ice for as long as possible. The samples were briefly mixed with pipetting immediately before 2.0uL of each sample was pipetted onto its respective microscope slide. The samples were covered with a glass cover and cemented to the slide using clear nail polish. The slides were kept out of the light until observation und a confocal scanning laser microscope.

Western blot analysis

Western blot analysis was carried out according to the protocol described in Santangelo et al. 2013. For the biotinylated methylguanine analog Western blot analysis, as primary antibody a mouse anti-Biotin antibody (Bethyl Laboratories) was used in a 1:6000 dilution. Detection of bands was carried out with an anti-mouse-HRP (horseradish peroxidase) secondary antibody.

Research and Progress

Markerless deletion of TK1971

In order to decrease background TK-MGMT activity, the native copy of TK1971 was deleted from the *T. kodakarensis* genome using a standard double recombination strategy utilized by the Reeve Lab in order to generate the *T. kodakarensis* deletion library (figure 2). The resulting deletion strain, Δ1971, was designated DJA7 (Table 1). The TK1971 deletion from *T. kodakarensis* strain TS559 was confirmed with PCR and

Southern blot analysis (figure 3B, 3C).

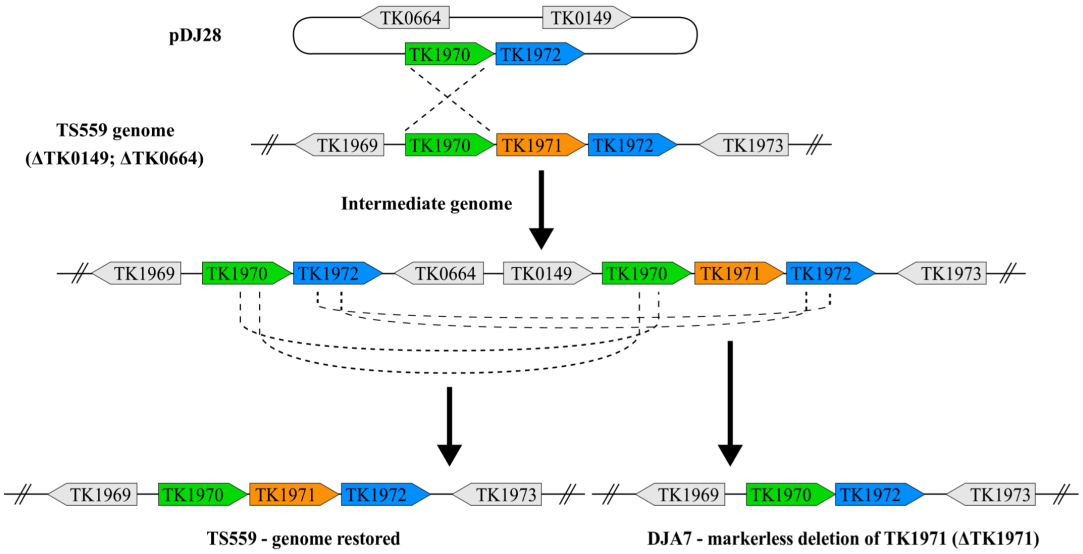

Figure 2. Markerless deletion of TK1971. The *T. kodakarensis* strain TS559 has two genes, TK0149 and TK0664, that have been deleted from its genome. TK1971 (orange) is initially present in the *T. kodakarensis* genome. Homologous recombination between the *T. kodakarensis* genome and the TK1971 deletion construct (pDJ28) results in the integration into the host chromosome yielding in two potential agmatine-prototrophic intermediate strains. pDJ28 contains regions of homology to the genes surrounding the TK1971 gene, TK1970 and TK1972 (blue and green), and dependent on recombination responsible for excision upon 6-methylpurine selection, either the host genome is restored (bottom left) or the desired deletion (DJA7) is generated (bottom right).

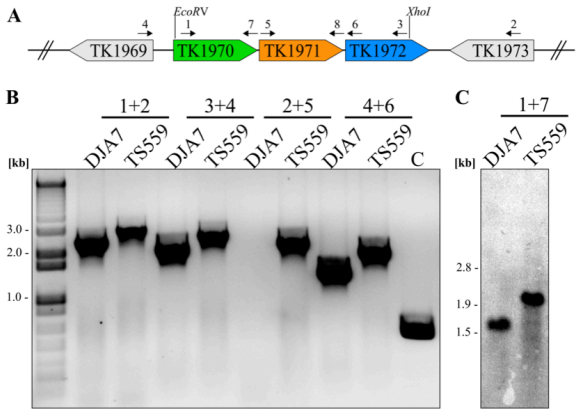

Figure 3. (A) *T. kodakarensis* strain TS559, with genes TK1970, TK1971, and TK1972 shown. Primers used for PCR analysis (labeled 1-8) are shown above genes. **(B)** PCR to confirm TK1971 deletion. Samples were separated on a 1.0% agarose gel with electrophoresis and stained with ethidium bromide. Smaller DNA sizes in strain DJA7 indicate the TK1971 gene, which is around 1kb, has been successfully deleted. **(C)** Southern blot analysis to confirm deletion of gene TK1971 in strain DJA7. gDNA from strains TS559 and DJA7, as well as a control, was digested with *EcoRV* and *XhoI*. Fragments were separated using electrophoresis with a 0.75% agarose gel, denatured, and transferred to a Zeta-Probe membrane (Bio-Rad). The membrane was probed with DIG-labeled PCR products from primers 1 and 7.

Reporter Construct Creation

The TK1971 gene was fused to a strong, constitutive promoter (P_{htmB}) to increase TK1971 expression and fluorescent signal in the reporter system, as well as a multiple cloning site in order to facilitate in-frame protein fusion. Multiple variations of this basic reporter construct exist. Protein tags, including N and C-terminal hemagglutins (HA) tags and an N-terminal Histidine (6xHis) tag, were fused to the TK1971 gene, in order to facilitate Western blotting and protein purification (figure 4).

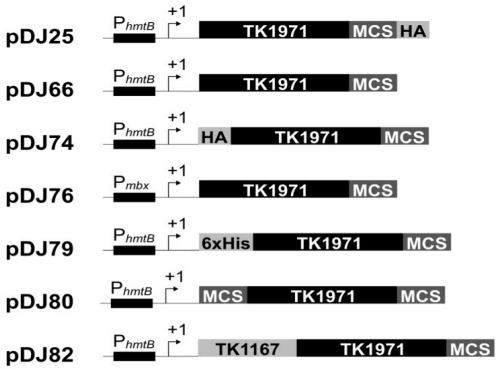

Figure 4. A schematic depiction of the TK1971 constructs in plasmid pTS543 are shown. TK1971 has been fused to a strong promoter, P_{htmB} (constitutive) or P_{mbx} (sulfur-dependent) and a multiple cloning site (MCS), and was cloned in to a standard cloning vector (pUC19). The transcription start site is denoted with +1, and the translation start codon is indicated by ATG. Some versions of the TK1971 reporter construct have an N or C terminal hemagglutins (HA) or histine (6xHis) tag, and one version has been fused to a 94 amino acid subunit of the *T. kodakarensis* RNA Polymerase (TK1167). See table 2 in supplementary figures.

PCR-based mutagenesis and cloning of TK1971 derivatives

Based on research for the SNAP-tag system developed by NEB, a series of mutations into TK1971 that are suspected to increase TK-MGMT's fluorescent activity as well as its affinity for the fluorescent methylguanine substrate were introduced using an inverse PCR strategy or site-directed mutagenesis on *E. coli* plasmid pUC19 (figure 1). The TK1971 reporter constructs were cut and ligated into a plasmid backbone capable of replicating in *T. kodakarensis* (pTS543) using standard digestion and ligation protocols. Once successfully cloned into plasmid pTS543, plasmids were transformed DJA7 (Table 1).

Growth Curve Experiments

To ensure the deletion of the native copy of TK1971 as well as the pTS543 based copy of the TK1971 gene did not interfere with normal *T. kodakarensis* growth, growth experiments on multiple strains of *T. kodakarensis*, TS559, DJA7, DJA8, and DJA9 (Table 1), were run until the strains reached a stationary growth phase, typically between 12 and 16 hours. Strains were grown in multiples of three, in both sulfur and pyruvate media. The OD_{600} of the *T. kodakarensis* cultures were taken every hour using a Bio-Rad spectrophotometer. The growth curve for all four strains indicate that neither deletion of the native copy of TK1971 nor the plasmid based TK1971 significantly affect *T. kodakarensis* growth (figure 5).

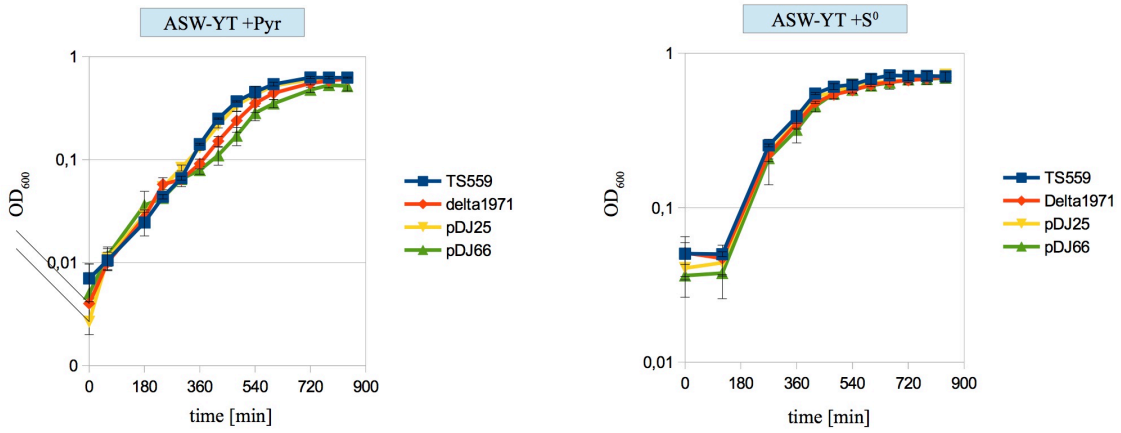

Figure 5. Growth behavior of selected *T. kodakarensis* strains in both pyruvate (left plot) and sulfur (right plot) media. Three strains *of T. kodakarensis*, TS559 (blue), DJA7 (Δ1971, orange), DJA8 (pDJ25, yellow), and DJA9 (pDJ66, green). *T. kodakarensis* strains were grown over the course of several hours. The optical density (OD_{600}) of the culture was taken every hour using a spectrometer.

Western blot analysis

To confirm the presence of the TK-MGMT enzyme in *T.* kodakarensis, TK-MGMT presence was visualized using Western blot experiments. The presence of TK-MGMT in *T. kodakarensis* strain DJA8 was detected using an anti-HA antibody, showing

a protein at 22 kDa, the size of the TK-MGMT protein (figure 6A). The deletion strain DJA7 was also tested with an anti-HA antibody, and no copies of the TK-MGMT protein appeared to be present (figure 6A).

Western blot analysis was also used to visualize the methyltransferase activity of TK-MGMT. As previously mentioned, multiple methylated guanine analogs are commercially available, including biotinylated methylguanine nucleotides. The MGMT protein is tagged with a biotin molecule upon the addition of the biotinylated nucleotides. TK-MGMT was incubated with the biotinylated methylguanine analogs for increasingly long time periods, 30 minutes, 60 minutes, and 180 minutes. Regardless of incubation time, TK-MGMT appeared to readily bind the biotinylated methyl group and showed a strong band at 22kDa (figure 6B). The consistency in signal, regardless of incubation time, could be because the methyltransferase reaction takes place quickly, and the TK-MGMT sample is labeled with biotin after 30 minutes. These results could also indicate that the substrate is the limiting factor in this methyltransferase reaction, and is depleted after 30 minutes.

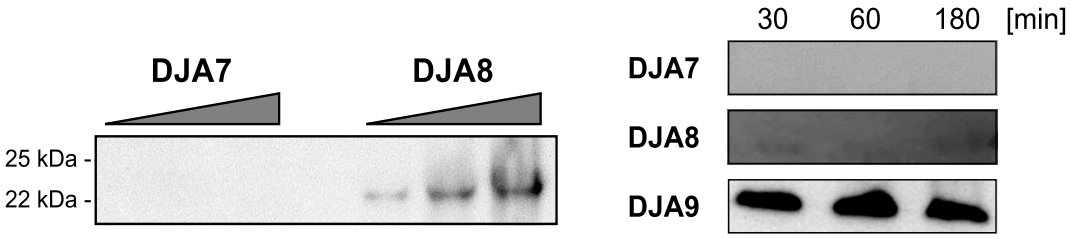

Figure 6. **(A)** Anti-HA Western blot analysis. Increasing concentrations (5, 10, 25 ug) of crude cell lysates of DJA7 (ΔTK1971) and DJA8 (DJA7 with pDJ25) were resolved on a 15% SDS-PAGE gel. The separated proteins were transferred to a PVDF membrane and probed with a rabbit antibody against the HA- epitope. Signal detection was performed using an anti-rabbit HRP (horseradish peroxidase) conjugate with ECL Plus as substrate. The Western shows a band at 22kDa, the size of the TK-MGMT protein. **(B)** Anti-biotin Western blot analysis. Strains DJA7, DJA8, and DJA9 were incubated with a biotinylated methylguanine analog for increasing time periods (30, 60, and 180 minutes, 25ug of crude cell lysate per lane) and the TK-MGMT proteins were resolved on a 15% SDS-PAGE gel. The separated proteins were transferred to a PVDF membrane and probed with a rabbit antibody against biotinylated TK-MGMT. Signal detection was performed using an anti-rabbit HRP (horseradish peroxidase) conjugate with ECL Plus as substrate.

CSLM results

The fluorescent activity of the reporter construct *in vivo* in *T. kodakarensis* was observed by introducing fluorescent methyguanine analogs into *T. kodakarensis* and observing the cells under a confocal scanning laser microscope. Multiple strains of *T. kodakarensis* have been observed thus far. Here we focus primarily on strains TS559, DJA7, DJA8, and DJA9 (Table 1, figure 7). CSLM images indicate that DJA8, which has a C-terminal HA-tag, is weakly and inconsistently fluorescently active. Increasing the concentration of fluorescent substrate during incubation did not change the fluorescent signal, and gave images with a high level of background fluorescence. Thus far, the TK1971 construct without any additional mutations or tags (DJA9) is the only construct that has shown consistently detectable fluorescence (figure 7). The fluorescent activity for this construct is easily detectable and reproducible. These results, in conjunction with the biotin Western blot results, indicate that TK-MGMT is binding specifically to the introduced methylguanine analogs. Ideally, the fluorescent signal would be stronger for TK1971's use as a reporter construct. The consistent fluorescence in strain DJA9 is promising for its use as a reporter gene system, particularly to monitor the strength of cis-regulatory elements.

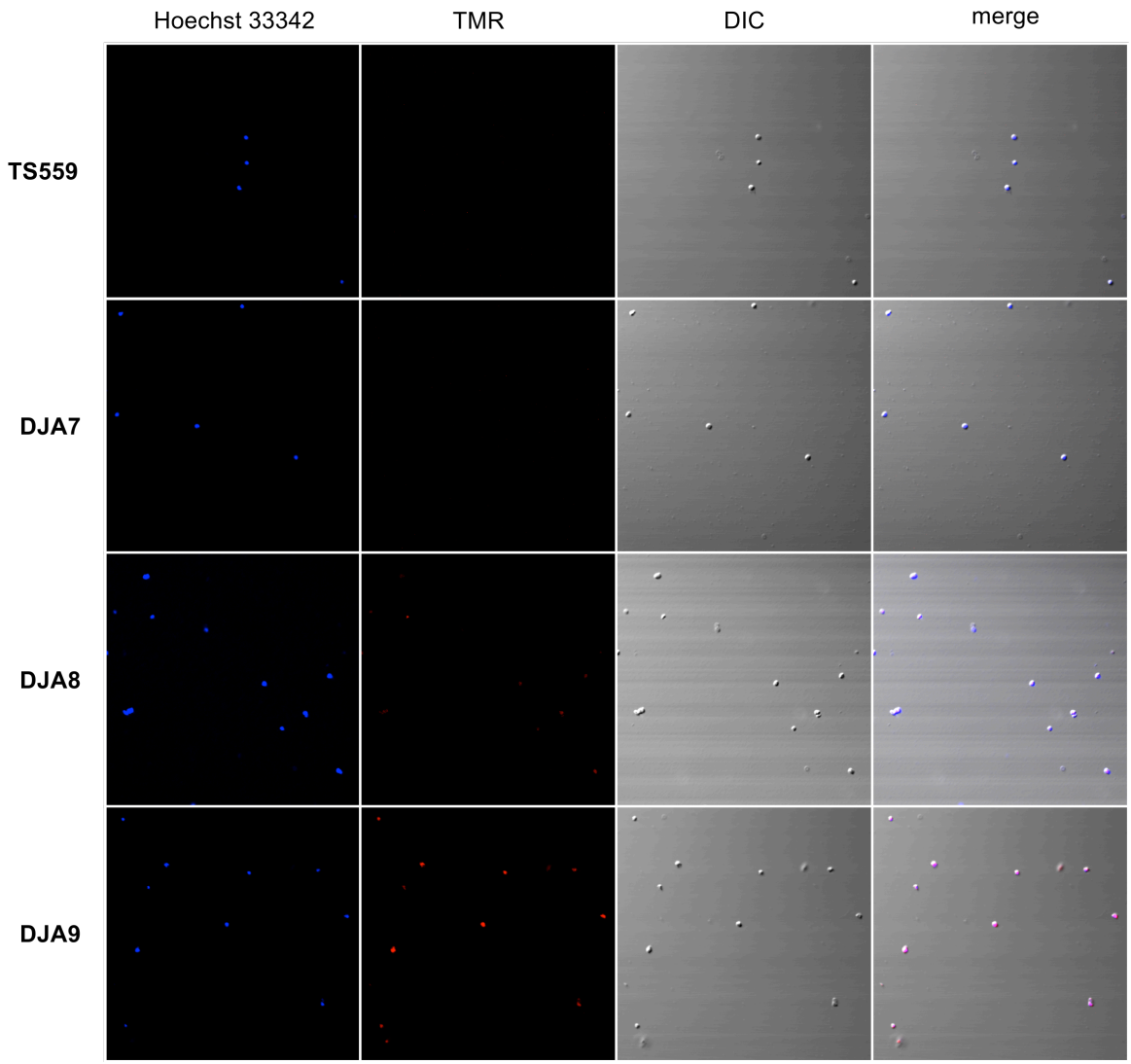

Figure 7. Microscopic analysis of selected *T. kodakarensis* strains 0.1 OD600 of *T. kodakarensis* cells were incubated in complemented ASW-YT medium containing 5 µM of TMR-Star for 1 h at 85°C. The cells were washed six times with ASW-YT and resuspended in 1 ml of medium containing 5 µM of Hoechst 33342, followed by a 30 min incubation at at 85°C. The cells were washed twice with 0.8x ASW and subjected to microscopic analysis using a Olympus FV1000 confocal laser scanning microscope (CLSM). Images were recorded using the Argon tube laser (458, 488, 515 nm) for the Hoechst 33342 and the HeNe laser (543 nm) for TMR-Star detection with a 60x oil immersion objective and 3x digital zoom. Differential interference contrast (DIC) was used to capture unlabeled cells. All images were acquired at identical microscopic settings.

Current Studies

Current experiments focus on fusing a small portion of the RNA polymerase protein (TK1167), designated rpoL, within the reading frame of the TK-MGMT protein in order to examine the strength of the reporter system for protein localization

experiments. To date, the rpoL fragment has been fused within the reading frame of the TK1971 construct, both with and without the C-terminal HA-tag. This TK1971:rpoL fusion has yet to be ligated into the pTS543 shuttle vector capable of replicating in *T. kodakrensis*. Future experiments will focus on transforming this construct into *T. kodakarensis* and quantifying the fluorescent activity of the TK-MGMT:rpoL fusion using CSLM. Additionally, there are TK1971 reporter constructs that have yet to be introduced and examined in *T. kodakarensis*. Future experiments will focus on introducing every variant of the reporter construct into *T. kodakarensis* and observing its fluorescent activity using CSLM. It is also possible that there are mutations not described in the literature that could significantly enhance the fluorescent activity of the TK-MGMT protein. A potential future experiment could focus on random mutagenesis of the TK1971 gene using error-prone PCR, and look for fluorescently active mutant cells after incubation with fluorescent methylguanine analogs. Fluorescence Activated Cell-Sorting allows for individual cells in a sample to be sorted out according to strength of fluorescent signal, so that the mutants that were the most fluorescently active could be isolated, and further examined.

Because the MGMT protein is a DNA repair enzyme, it stands to reason that TK-MGMT production could be quantified measuring *T. kodakarensis'* response to DNA damaging agents. Multiple strains *T. kodakarensis*, both with and without the TK-MGMT based reporter system, could be exposed to DNA damaging agents that methylate specifically guanine nucleotides at the O^6 position. Cells with abundant levels of TK-MGMT should recover quickly after being exposed to the DNA methylating agents because the MGMT protein specifically repairs guanine nucleotides that have a methyl

group at the O^6 position. Quantifying TK-MGMT's response to DNA damaging agents could provide valuable insight into the amount of TK-MGMT being produced in *T. kodakarensis* and the efficiency of its enzymatic activity.

Conclusion

The version of the TK1971 reporter gene system without any additional mutations and without an N or C-terminal tag, DJA9, has shown the most fluorescent activity after being incubated with the fluorescent methylguanine analogs. The addition of a C-terminal HA-tag corresponds to a significant decrease is fluorescent signal (figure 7)**.** The enzymatic activity of reporter construct pDJ74, which contains an N-terminal HA-tag, has also been examined with CSLM after being incubated with fluorescent methylguanine analogs, with similar results. At first glance, this does not bode well for TK-MGMT's use as a reporter enzyme. The HA-tag is only nine amino acids long, and based on CSLM results, appears to significantly interfere with the structure of the TK-MGMT protein and inhibit its enzymatic activity. The addition of the HA-tag could interfere with the TK-MGMT enzyme's ability to fold properly after it is translated, thus interfering with TK-MGMT's catalytic site for the methyltransferase reaction. It is possible that the HA-tag is masking the effects of the mutations that have been introduced into TK1971 in order to increase TK-MGMT's fluorescent activity. Reporter construct pDJ79 has an N-terminal Histidine (6xHis) tag. Future experiments will focus on this construct's ability to bind and remove methylguanine analogs, to see if TK-MGMT is active when attached to a different small protein tag.

In order to be useful for protein localization experiments, the TK-MGMT enzyme needs to be functional when it is fused with much larger proteins. Future experiments

focus on the activity of TK-MGMT when it is fused to a larger protein, a subunit of the *T. kodakarensis* RNA Polymerase (TK1167). These experiments will likely examine the effects of fusing a protein of interest to both the N and C-terminal of the TK-MGMT enzyme. Future experiments should also focus on the use of the TK-MGMT reporter gene system to observe promoter strength by exchanging the P_{htmB} promoter for a different promoter of choice and comparing the strength of the resulting fluorescent TK-MGMT signal to the original construct.

Western blot analysis detecting the presence of biotinylated TK-MGMT (figure 6B) demonstrate that the TK-MGMT enzyme present in strain DJA9 is active and capable of successfully removing and permanently binding tagged methyl groups from methylated guanine analogs. These results correlate well with the CSLM imaging results (figure 7), suggesting that the microscopy results are due to legitimate and specific TK-MGMT binding. This specificity of binding and strength and consistency of fluorescent signal in this TK1971 construct shows significant promise for use as a reporter system.

The fluorescent signal from the DJA9 reporter construct is consistently visible with CSLM, and the strength of the signal is reproducible. It has potential to be useful for studies on promoter strength, and potentially protein localization. However, there are also combinations of mutations introduced into the TK1971 gene that have yet to be observed with CSLM. Any one of these mutations or combinations of mutations has the potential to be more active with the fluorescent methylguanine analogs than the basic DJA9 construct, but they have yet to be examined *in vivo* in *T. kodakarensis*. Once in place, the TK1971 based reporter gene system will serve as an indispensible tool to visualize gene expression as well as protein localization in *T. kodakarensis* and other

hyperthermophilic microorganisms.

Supplementary figures

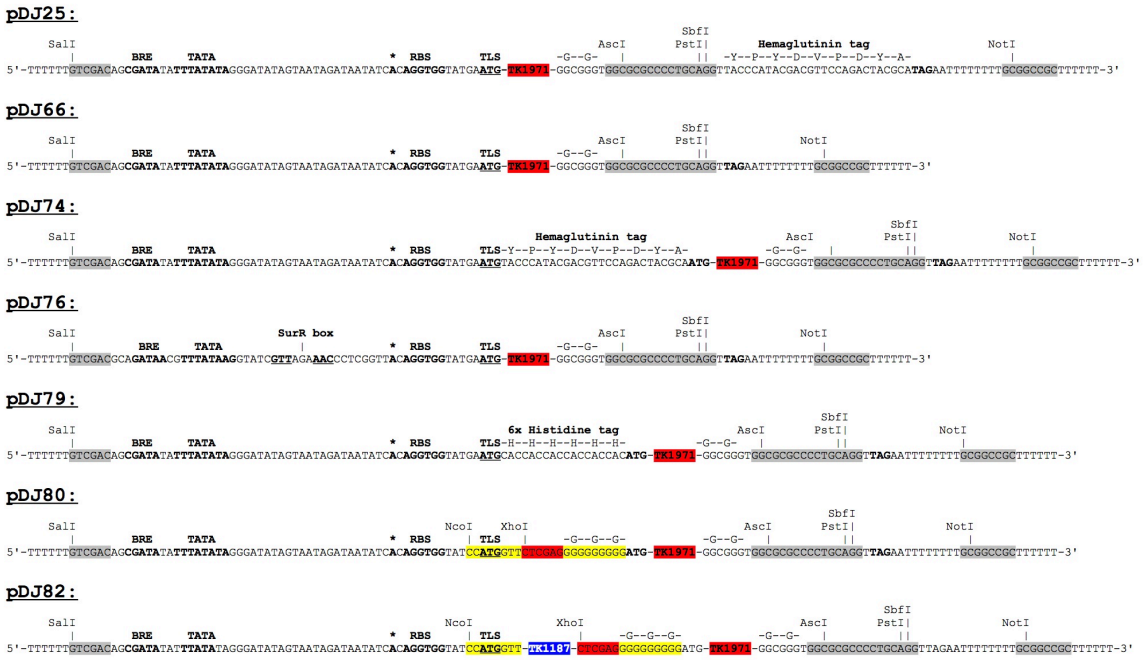

Figure S1. A more detailed depiction of the TK1971 reporter constructs, with restriction enzyme cut sites, BRE and TATA boxes, the ribosomal binding site (RBS), translation start site (TLS), and the master regulatory gene for sulfur metabolism in *T. kodakarensis* (SurR box) highlighted and labeled.

Strain	Plasmid	Mutations	Promoter	Tag	Backbone
OSU0017	pDJ14	wild-type	*P*hmtB	NO tag	pUC19
OSU0018	pDJ15	wild-type	*P*hmtB	C-HA	pUC19
OSU0071	pDJ68	Yellow, blue (figure 1)	*P*hmtB	C-HA	pUC19
DJA0008	pDJ25	wild-type	*P*hmtB	C-HA	pTS543
DJA0009	pDJ66	wild-type	*P*hmtB	NO tag	pTS543
DJA0022	pDJ74	wild-type	*P*hmtB	N-HA	pTS543
DJA0024	pDJ76	wild-type	*P*mbx	NO tag	pTS543
DJA0027	pDJ79	wild-type	*P*hmtB	N-His	pTS543
DJA0030	pDJ82	wild-type, +rpoL fusion	*P*hmtB	NO tag	pTS543

Table 2. Table outlining the plasmids used to generate strains for TK1971 reporter construct.

1. Woese CR, Fox GE. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc. Natl. Acad. Sci. USA 74(11):5088-5090.

2. Woese, CR, Kandler O, Wheelis ML. 1990. Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria and Eucarya. Proc Natl Acad Sci USA. 87(12): 4576-4579.

3. Brochier-Armanet C, Forterre P, Gribaldo S. 2011. Phylogeny and evolution of the Archaea: one hundred genomes later. Curr. Opin. Microbiol. 14(3):274-81.

4. Garrett RA, Klenk HP. 2007. *Archaea. Evolution, Physiology, and Molecular Biology.* Blackwell Publishing, Oxford, UK.

5. Ishino Y, Ishino S. 2012. Rapid progress of DNA replication studies in Archaea, the third domain of life. Sci China Life Sci. 55(5): 386-403.

6. Benelli D, Londei P. 2011. Translation initiation in Archaea: conserved and domain-specific features. Biochem Soc Trans. 39(1): 89-93.

7. La Teana A, Benelli D, Londei P, Bläsi U. 2013. Translation initiation in the crenarchaeon Sulfolobus solfataricus: eukaryotic features but bacterial route. Biochem Soc Trans. 41(1):350-355.

8. Grohmann D, Werner F. 2011. Recent advances in the understanding of archaeal transcription. Current Opinions in Microbiology. 14(3): 328-334.

9. Bell S, Jackson S. 1998. Transcription and Translation in Archaea: a mosaic of eukaryal and bacterial features. Trends in Microbiology. 6(6), 222-228.

10. Atomi H, Sato T, Kanai T. 2011. Applications of hyperthermophiles and their enzymes. Current Opinion in Biotechnology. 22 (5): 618-26.

11. Egorova K, Antranikian G. 2005. Industrial relevance of thermophilic Archaea. Curr. Opin. Microbiol. 8(6): 649-655

12. Sato T, Fukui T, Atomi H, Imanaka T. 2003. Targeted gene disruption by homologous recombination in the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. J. Bacteriol. 185(1):210-220.

13. Atomi H, Fukui T, Kanai T, Morikawa M, Imanaka T. October 2004. Description of *Thermococcus kodakaraensis* sp. nov., a well studied hyperthermophilic archaeon previously reported as *Pyrococcus* sp. KOD1. Archaea. 1(4): 263–267.

14. Leigh JA, Albers S, Atomi H, Allers T. December 2010. Model organisms for

genetics in the domain Archaea: methanogens, halophiles, *Thermococcales* and *Sulfobales.* FEMS Microbiol Rev. 35(4): 577-608.

15. Cavicchioli, R. 2007. *Archaea. Molecular and Cellular Biology.* ASM Press. Washington, DC, USA.

16. Santangelo, T, Cubonova L, Reeve J. May 2008. Shuttle Vector Expression in *Thermococcus kodakakrensis:* Contribution of *cis* Elements to Protein Synthesis in a Hyperthermophilic Archaeon. Appl Environ Microbiol. 74(10): 3099–3104

17. Santangelo TJ, Cubonova L, Reeve JN. 2010. *Thermococcus kodakarensis* genetics: TK1827-encoded beta-glycosidase, new positive-selection protocol, and targeted and repetitive deletion technology. Appl. Environ. Microbiol. 76(4): 1044-1052.

18. SNAP-tag® Technologies: Novel Tools to Study Protein Function https://www.neb.com/tools-and-resources/feature-articles/snap-tag-technologies-novel-tools-to-study-protein-function.

19. Spratt T, Wu JD, Levy DE, Kanugula S, Pegg A. 1999. Reaction and Binding of Oligodeoxynucleotides with Wild-Type and Mutant Human O^6-Alkylguanine-DNA Alkyltransferase. Biochemistry. 38(21): 6801-6806.

20. Gronemeyer T, Chidley C, Juillerat A, Heinis C, Pick H, Johnsson K. 2006. Directed Evolution of O^6-Alkylguanine-DNA Alkeyltransferase for applications in protein labeling. Protein Engineering, Design & Selection. 19(7): 313-317.

21. Hashimoto, H., Inouel, T., Nishioka, M., Fujiwara, S., Takagi, M., & Kail, Y. (1999). Hyperthermostable Protein Structure Maintained by Intra and Inter-helix Ion-pairs in Archaeal O6-Methylguanine-DNA Methyltransferase. *Journal of Molecular Biology*, *292*, 707–716.

22. Shiraki, K., Nishikori, S., Fujiwara, S., Hashimoto, H., Kai, Y., Takagi, M., & Imanaka, T. (2001). Comparative analyses of the conformational stability of a hyperthermophilic protein and its mesophilic counterpart. *European Journal of Biochemistry*, *268*(15), 4144–4150.

23. Perugino, G., Vettone, A., Illiano, G., Valenti, A., Ferrara, M. C., Rossi, M., & Ciaramella, M. (2012). Activity and regulation of archaeal DNA alkyltransferase: Conserved protein involved in repair of DNA alkylation damage. *Journal of Biological Chemistry*, *287*(6), 4222–4231.

24. Kanugula, S., & Pegg, A. E. (2003). Alkylation damage repair protein O6-alkylguanine-DNA alkyltransferase from the hyperthermophiles Aquifex aeolicus and Archaeoglobus fulgidus. *The Biochemical Journal*, *375*(Pt 2), 449–455.

25. Pan, M., Santangelo, T. J., Čuboňová, Ľ., Li, Z., Metangmo, H., Ladner, J., … Kelman, Z. (2013a). Thermococcus kodakarensis has two functional PCNA homologs but only one is required for viability. *Extremophiles*, *17*(3), 453–461.

26. Santangelo, T., Čuboňová, L., James, C. L., & Reeve, J. N. (2007). TFB1 or TFB2 is sufficient for Thermococcus kodakaraensis viability and for basal transcription in vitro. *Journal of Molecular Biology*, *367*(2), 344–257.

27. Juillerat, A., Gronemeyer, T., Keppler, A., Gendreizig, S., Pick, H., Vogel, H., & Johnson, K. (2003). Directed Evolution of O6 -Alkylguanine-DNA Alkyltransferase for Efficient Labeling of Fusion Proteins with Small Molecules In Vivo. *Chemistry & Biology*, *13*(15), 1317–1323.

28. Gronemeyer, T., Chidley, C., Juillerat, A., Heinis, C., & Johnsson, K. (2006). Directed evolution of O6-alkylguanine-DNA alkyltransferase for applications in protein labeling. *Protein Engineering, Design and Selection*, *19*(7), 309–316.

29. Keppler, A., Kindermann, M., Gendreizig, S., Pick, H., Vogel, H., & Johnsson, K. (2004). Labeling of fusion proteins of O6-alkylguanine-DNA alkyltransferase with small molecules in vivo and in vitro. *Methods*, *32*(4), 437–444.

30. Juillerat, A., Heinis, C., Sielaff, I., Barnikow, J., Jaccard, H., Kunz, B., … Johnsson, K. (2005). Engineering substrate specificity of O6-alkylguanine-DNA alkyltransferase for specific protein labeling in living cells. *ChemBioChem*, *6*(7), 1263–1269.

31. Santangelo, T. J., Čuboňová, L., & Reeve, J. N. (2010). Thermococcus kodakarensis Genetics: Tk1827-encoded β-glycosidase, new positive-selection protocol, and targeted and repetitive deletion technology. *Applied and Environmental Microbiology*, *76*(4), 1044–1052.